BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37141
doi: 10.3897/biss.3.37141

OPEN ACCESS

Conference Abstract

ALICE, MALICE and VILE: High throughput insect specimen digitisation using angled imaging techniques

Steen Dupont‡, Benjamin W Price‡

‡ Natural History Museum, London, United Kingdom

Corresponding author: Steen Dupont (steen.dupont@nhm.ac.uk)

Received: 11 Jun 2019 | Published: 13 Jun 2019

Citation: Dupont S, Price BW (2019) ALICE, MALICE and VILE: High throughput insect specimen digitisation using angled imaging techniques. Biodiversity Information Science and Standards 3: e37141.
https://doi.org/10.3897/biss.3.37141

Abstract

The world's natural history collections contain at least 2 billion specimens (Ariño 2010), representing a unique data source for answering fundamental scientific questions about ecological, evolutionary, and geological processes. Unlocking this treasure trove of data, stored in thousands of museum drawers and cabinets, is crucial to help map a sustainable future for ourselves and the natural systems on which we depend. The rate-limiting steps in the digitisation of natural history collections often involve specimen handling, due to their fragile nature. Insects comprise the single largest collection type in the Natural History Museum, London (NHM) and in many other collections, reflecting their global diversity and multiplicity. The NHM pinned insect collection, estimated at 25 million specimens, will take over 700 person years to digitise at current rates (Price et al. 2018: estimated from Blagoderov et al. 2017).

In order to ramp up digitisation, we have developed ALICE for Angled Label Image Capture and Extraction from pinned insects. This multi-camera setup (Fig. 1) and associated software processing pipeline, enables primary data capture from angled images, without removal of the labels from the specimen pin. As a result ALICE enables a single user to sustainably digitise (add a catalogue label, image and prepare images for database import) over 800 specimens per day (Price et al. 2018), allowing us to digitally unlock large parts

of the insect collection (e.g., Hymenoptera, Diptera, Coleoptera) at up to seven times the previous rate.

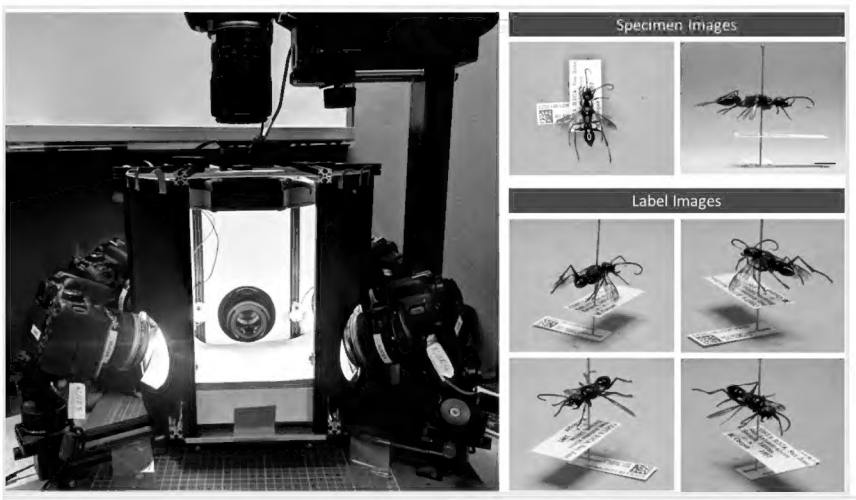

Figure 1.

ALICE multi-camera setup *in situ* and examples of specimen overview and label images captured by each of the cameras. Note: example images are cropped for display.

We are continuing to refine hardware approaches to reduce specimen handling and extract data, for both human and machine interpretation, from labels without removing them from the object. More recently we are also trialing multiple mirrors in our Mirror Angled Label Image Capture Equipment (MALICE) (Fig. 2) or a rotating stage for our Vial Image Label Extraction (VILE) (Fig. 3) aimed at spirit-preserved specimens housed in vials. In this talk, we will outline the current approaches in use at the Natural History Museum, next generation prototypes, and challenges that need to be addressed before these techniques can be fully optimized.

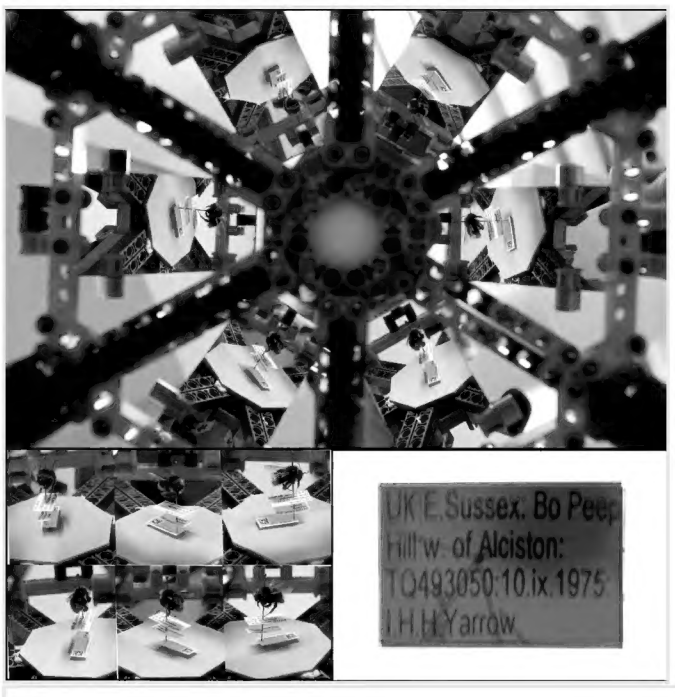

Figure 2.

The image output of an early mirrored ALICE prototype including manual crop and deskew of the six specimen views and a label stack consisting of the six superimposed extracted label crops.

Figure 3.

Overview of the rotating stage setup for vial label imaging, with five images of an example vial, combined into a single overview image enabling transcription of the visible label data.

Keywords

digitisation, pinned specimens, spirit preserved specimens, label extraction

Presenting author

Steen Dupont

Presented at

Biodiversity_Next 2019

Acknowledgements

We thank Louise Allan, Laura Balcells, Vladimir Blagoderov, Alice Butcher, James Durrant, Pieter Holtzhausen, Sameh Sharif, Steve Suttle, Glen Moor and Phaedra Kokkini for their contributions to developing and testing the workflows, hardware and associated software. These projects have received support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777483 "ICEDIG"; the SYNTHESYS Project, http://www.synthesys.info/, which is financed by European Community Research Infrastructure Action under the FP7 Integrating Activities Programme

(Grant agreement number 312253); and the Natural History Museum's Digital Collections Programme.

References

- Ariño A (2010) Approaches to estimating the universe of natural history collections data. Biodiversity Informatics 7 (2). https://doi.org/10.17161/bi.v7i2.3991
- Blagoderov V, Penn M, Sadka M, Hine A, Brooks S, Siebert D, Sleep C, Cafferty S, Cane E, Martin G, Toloni F, Wing P, Chainey J, Duffell L, Huxley R, Ledger S, McLaughlin C, Mazzetta G, Perera J, Crowther R, Douglas L, Durant J, Scialabba E, Honey M, Huertas B, Howard T, Carter V, Albuquerque S, Paterson G, Kitching I (2017) iCollections methodology: workflow, results and lessons learned. Biodiversity Data Journal 5: e21277. https://doi.org/10.3897/bdj.5.e21277
- Price BW, Dupont S, Allan EL, Blagoderov V, Butcher AJ, Durrant J, Holtzhausen P, Kokkini P, Livermore L, Hardy H, Smith V (2018) ALICE: Angled Label Image Capture and Extraction for high throughput insect specimen digitisation. SF Preprints https://doi.org/10.31219/osf.io/s2p73